



« TABOU SEARCH » METHODE DANS LA GENERATION ASSISTEE SUR ORDINATEUR DES COMPOSES MACROMOLECULAIRES ♦

Teodora Rusu¹, Oana Marlena Gogan²

¹*Institut de Chimie Macromoléculaire ‘P. Poni’, 6600 Jassy, Roumanie,
teia@tuiasi.ro*

²*S. C. ARTINFO SRL, Iași – Division Roumaine du ARTI INFOMATICHE,
ITALIE, ogogan@cs.tuiasi.ro*

Abstract: The Tabou method is a strong heuristic method giving very good results for monocriterial problems. The numerical results obtained in multiobjective problems are presented, and Tabou method is compared with other known methods (genetic algorithms). The genetic algorithms are already used in the computerized optimization analysis of the macromolecular products synthesis. The generation of the initial population may be done by random creation of a fixed number of chromosomes, imposing the condition that each chromosome is a valid solution of the desired macromolecular structure. The evaluation of a population is done by assigning to each individual an evaluation function. The work presents an experimental study of using mixed genetic – Tabou algorithms applied in a multicriteria context.

Keywords: *Tabou method, genetic algorithms, macromolecular compound synthesis*

♦ Paper presented at COFrRoCA 2006: Quatrième Colloque Franco-Roumain de Chimie Appliquée, 28 June – 2 July, Clermont-Ferrand, France

INTRODUCTION

Depuis toujours, les chercheurs ont tenté de résoudre les problèmes N P – difficiles le plus efficacement possible. Pendant longtemps, la recherche s'est orientée vers la proposition d'algorithmes exacts pour des cas particuliers polynomiaux. Ensuite, l'apparition des heuristiques a permis de trouver des solutions en général de bonne qualité pour résoudre les problèmes. En même temps, les méthodes de type « séparation et évaluation » ont aidé à résoudre des problèmes de manière optimale, mais souvent pour des instances de petite taille. Lorsque les premières metaheuristiques apparaissent, beaucoup de chercheurs se sont lancés dans l'utilisation de ces méthodes. Cela a conduit à une avancée importante pour la résolution pratique de nombreux problèmes [1,2]. Toutes les metaheuristiques s'appuient sur un équilibre entre l'intensification de la recherche et la diversification de celle-ci. D'un côté, l'intensification permet de rechercher des solutions de plus grande qualité en s'appuyant sur les solutions déjà trouvées et de l'autre, la diversification en place des stratégies qui permettent d'explorer un plus grand espace de solutions et d'échapper des minima locaux.

Dans notre cas on pressant une stratégie basée sur la théorie génétique pour l'identification ou l'optimisation des structures macromoléculaires avec des propriétés impose. Cette approche n'exige pas la synthèse et l'évaluation des bibliothèques énormes. Au lieu de cela elle comporte les générations itératives de plus petits ensembles témoin, qui sont analysés, et du "expérimentation"; la réponse biologique déterminée est employée comme une entrée pour GA pour trouver rapidement de meilleurs solution. Le GA décrit ici a été modifié par une Tabou Search algorithme. Les résultats prouvent que chaque génération successive a fourni de meilleures propriétés. Ces avancées sont démontrées à l'aide d'une étude de cas de conception de polymère rapporte à la *fitness function* du système étudié.

METHODES D'ANALYSE

Algorithme génétique (GA)

Les algorithmes génétiques sont basés sur la théorie de l'évolution de Darwin [1]. Une bonne solution à un problème est trouvée par évolution des solutions les plus adaptées. Le principe de base est le suivant (figure 1):

1. Initialement une population (un ensemble de solutions) est choisie au hasard.
2. Chaque individu (ou chromosome s'il n'y a qu'un seul chromosome par individu) voit son indice d'adaptation (fitness) évalué.
3. Une nouvelle population est créée pour la génération suivante en réitérant les étapes suivantes jusqu'à ce que celle-ci soit complète :
 - a. Sélection de 2 parents de la population dépendant de leur fitness.
 - b. Croisement (Crossover) d'après un taux de croisement. S'il y a croisement, 2 enfants sont produits, sinon les parents sont conservés tel quel dans la génération suivante. Le taux de croisement est en général élevé (70%).

- c. Mutation d'un individu par modification d'un locus (position dans le chromosome) de son chromosome d'après un taux de mutation (en général très faible 1%).
4. Cette nouvelle population remplace l'ancienne.
 5. L'algorithme s'arrête si le critère d'arrêt est atteint et renvoie la meilleure solution (le meilleur individu) de la génération sinon l'algorithme reprend en à l'évaluation de la fitness.

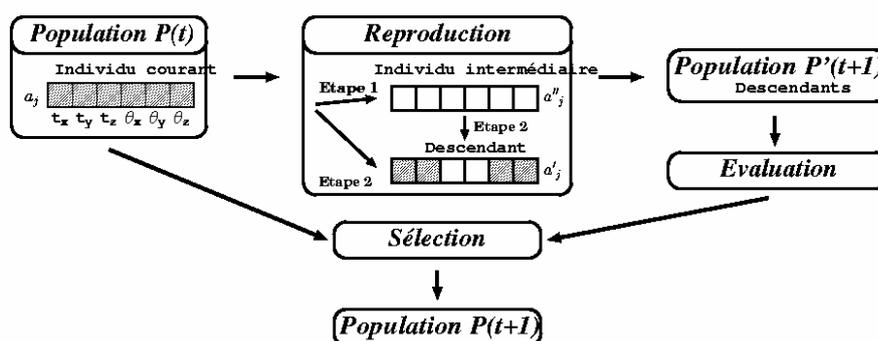


Figure 1. Schéma illustrant le principe de l'algorithme génétique

Dans cet algorithme, la sélection et le croisement permettent de converger rapidement vers une bonne solution et les mutations permettent d'éviter les maxima locaux (recherche dans l'espace d'états). Il est donc important et nécessaire que les mutations figurent dans l'algorithme pour éviter d'aboutir à une stagnation médiocre.

Tabu Search Method (Recherche Tabou) (TS)

La recherche Tabou est une métaheuristique d'optimisation [3]. Cette méthode est une métaheuristique itérative qualifiée de recherche locale au sens large. Le principe de la recherche Tabou est d'explorer le voisinage et choisir la position dans ce voisinage qui minimise la fonction objectif à partir d'une position donnée. Lorsque tous les points du voisinage ont une valeur plus élevée, cette opération peut conduire à augmenter la valeur de la fonction. On peut ainsi repérer les minima locaux (figure 2). Pour éviter à l'étape suivante de retomber dans un minima local, il faut stocker en mémoire les dernières positions explorées d'où le nom « tabou ». Il a été démontré pour la recherche Tabou que la convergence existe dans des conditions strictes rarement mis en pratique. On notera qu'il existe de nombreuses variantes aussi bien au niveau de la définition du voisinage que de la façon de gérer la mémoire. Pour mettre en place un algorithme fonctionnel et adapté au problème de la chimie macromoléculaire, nous avons dû faire des choix.

Nombre de chromosomes :

Un seul chromosome est nécessaire pour représenter l'ensemble de la solution. Le découpage un chromosome pour une tournée est particulièrement faux dans le sens où le nombre de tournées varie d'une solution à l'autre et un gène d'un chromosome peut très

bien se retrouver dans un autre. Ce chromosome est composé d'un ensemble ordonné de gènes.

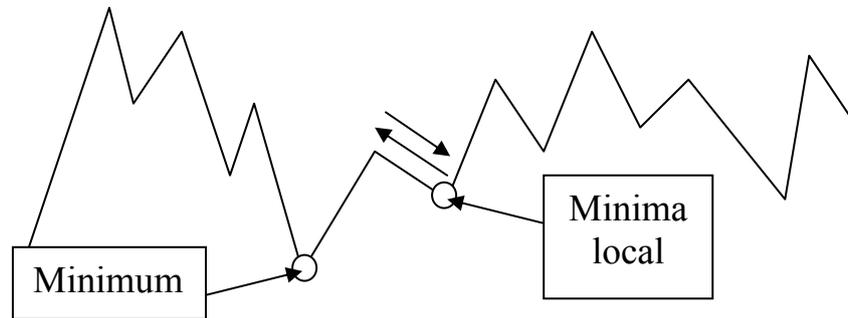


Figure 2.

Codage des gènes :

Pour la composition des gènes, plusieurs solutions étaient possibles :

- un gène est composé d'une tâche qui correspond à une voirie ramassable
- un gène est composé d'une tournée qui correspond à un ensemble de voiries ramassables
- un gène est composé d'une tâche qui est de type simple et qui permet d'identifier de manière unique une voirie ramassable

La première solution a été retenue, elle est relativement simple à mettre en œuvre même si elle implique un découpage ultérieur en tournée. La deuxième solution a été abandonnée car elle ne permet pas d'effectuer de bonnes mutations au sens génétique du terme. En effet des mutations comme swap ou mauve n'a aucun intérêt car elles ne font que déplacer et échanger des tournées au sein de la solution. De plus la diversité serait fortement réduite. La troisième solution est envisageable, elle sera peut-être implantée dans un avenir proche pour voir s'il y a un gain en temps de calcul.

Fitness et tournée :

Le codage des gènes choisi nous amène au problème d'évaluation du coût de la solution, la solution étant composée de tournées et non de voiries ramassables. Pour l'instant, un découpage séquentiel en tournée est réalisé lors de l'évaluation de l'individu (fitness) et le coût est calculé au niveau de la notation. La fitness calcule le coût de la solution, il faut donc conserver les individus de plus faible fitness.

Gestion des clones :

Les clones sont interdits dans la population, un clone est défini comme un individu dont la solution est identique (coût, ordonnancement des voiries). Lorsqu'un clone est détecté, l'individu est recréé au hasard. Si l'individu persiste à être un clone pendant un certain nombre d'itérations, il est supprimé définitivement de la population (ce qui n'arrive que si le nombre d'individus initial est trop important par rapport au problème à traiter). Une deuxième solution pour gérer les clones est envisagée : il s'agirait de n'éliminer que les individus de coût identique. Enfin une troisième possibilité est l'autorisation des clones dans la population sans aucun contrôle. Les solutions sont ordonnées par coût en temps de calcul, ainsi la première est très gourmande tandis que la dernière ne coûte rien.

Améliorations successives :

La mutation dite de recherche locale donne de bons résultats mais ralentit fortement l'algorithme. Pour pallier en partie à ce problème, le nombre d'améliorations possibles peut être limité : ainsi un chromosome n'essaie pas toutes les combinaisons jusqu'à arriver à une stagnation mais seulement un certain nombre. Les tests montrent que les calculs sont bien plus rapides, pour un résultat équivalent (figure 3).

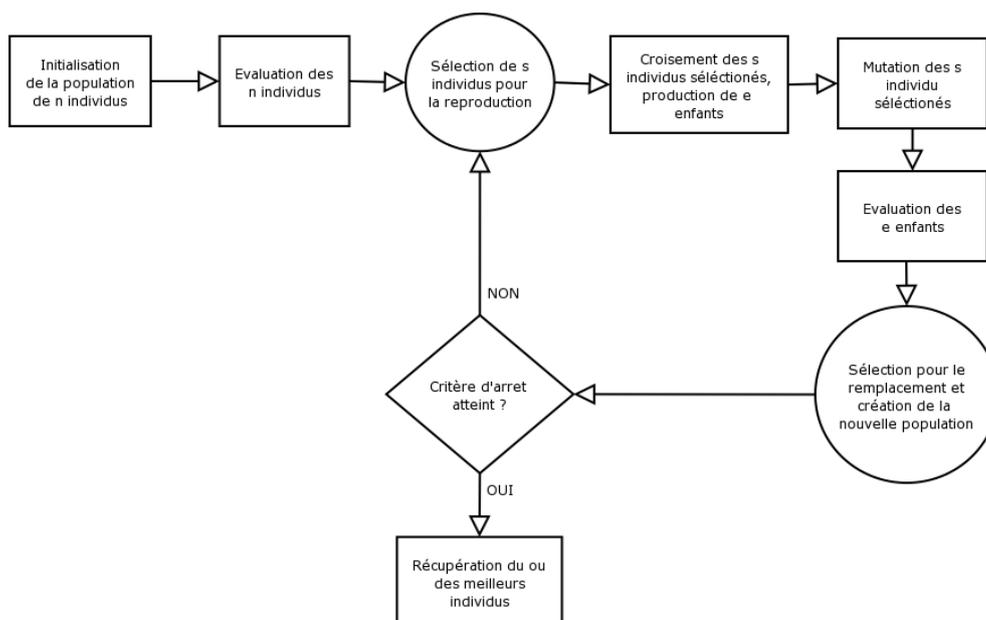


Figure 3. Diagramme d'état des algorithmes GA-TS

RESULTATS ET DISCUSSIONS

Dans cet travaux on pressant l'étude du synthèse optimise a l'ordinateur pour de copolymères contenant de séquence du polysyloxane, sur l'aspect du rate de succès du méthode propose [4, 5]. Dans ces étude on a considère les suivants oligosyloxanes fonctionnels (tableau 1).

Pour évaluer à quel point une molécule de candidat satisfait les propriétés et les contraintes désirées de cible, on a besoin d'une fonction de forme physique qui renvoie une "forme physique" numérique simple ou la "figure du mérite". La fonction de forme physique est problème dépendant et dépend de la nature des contraintes de propriété. Une fonction "gaussienne" pesée qui mesure la distance des propriétés de conception de candidat des valeurs à atteindre désirées courantes est utilisée (Eqn. 1) dans notre approche [6]. Nous cherchons à équiper des structures de copolymère n = 4 groups fonctionale. C'est maintenant deux paramètre, un j et ψ_s a ajuste en utilisant la fonction des erreur du moindres carrés

$$E = \int_{\psi_1}^{\psi_2} [\overline{K_c} - \overline{K_B}]^2 d\psi = E(j, \psi_s) \tag{1}$$

où la gamme de l'intégration resté toujours d'être choisie. Nous avons employé $\psi_1 = 1$. On a évite le point du ajustage ψ_s . Le point de commutateur de sorte que soit inclus

dans la gamme de l'ajustement, et l'ajustage de précision évite de rapprocher K_c avec une constante. Le tableau 2 récapitule les valeurs adaptées de j et de ψ_s pour différents tests.

Tableau 1. Oligosiloxanes fonctionnels

Oligosiloxanes fonctionnels	
1.	$\begin{array}{c} \text{CH}_3 \\ \\ \text{---}(\text{Si-O})_4\text{---} \\ \\ \text{CH}_3 \end{array} + \text{H}_2\text{C} = \text{CH} \begin{array}{c} \text{CH}_3 \text{ CH}_3 \\ \quad \\ \text{---Si-O---Si---} \\ \quad \\ \text{CH}_3 \text{ CH}_3 \end{array} = \text{CH}_2 \longrightarrow \text{H}_2\text{C} = \text{CH} \begin{array}{c} \text{CH}_3 \text{ CH}_3 \\ \quad \\ \text{---}(\text{Si-O})\text{---Si---} \\ \quad \\ \text{CH}_3 \text{ CH}_3 \end{array} = \text{CH}_2$ <p>(D₄) (DVTMDS)</p> <p>α, ω - divinil polysiloxane</p>
2.	$\begin{array}{c} \text{CH}_3 \\ \\ \text{---}(\text{Si-O})_4\text{---} \\ \\ \text{CH}_3 \end{array} + \text{H} \begin{array}{c} \text{CH}_3 \text{ CH}_3 \\ \quad \\ \text{---Si-O---Si---} \\ \quad \\ \text{CH}_3 \text{ CH}_3 \end{array} \text{H} \longrightarrow \text{H} \begin{array}{c} \text{CH}_3 \text{ CH}_3 \\ \quad \\ \text{---}(\text{Si-O})_n\text{---} \\ \quad \\ \text{CH}_3 \text{ CH}_3 \end{array} \text{H}$ <p>Si - H final</p>
3.	$\begin{array}{c} \text{CH}_3 \\ \\ \text{---O---Si---H} \\ \\ \text{CH}_3 \end{array} + \text{CH}_2 = \text{CH} \text{---} \text{CH}_2 \text{---} \text{OH} \longrightarrow \begin{array}{c} \text{CH}_3 \\ \\ \text{---O---Si---R---OH} \\ \\ \text{CH}_3 \end{array}$ <p>$\text{R} = \text{---}(\text{CH}_2)_2\text{---}$ sau $\begin{array}{c} \text{CH}_3 \\ \\ \text{---CH---} \\ \\ \text{CH}_2 \end{array}$</p> <p>Si - OH final</p>
4.	$\begin{array}{c} \text{CH}_3 \\ \\ \text{O---Si---R---O---OC---} \\ \\ \text{CH}_3 \end{array} \text{---} (\text{CH}_2)_2 \text{---} \begin{array}{c} \text{CH}_3 \\ \\ \text{C} \\ \\ \text{CN} \end{array} \text{---} \text{N} = \text{N} \text{---} \begin{array}{c} \text{CH}_3 \\ \\ \text{C} \\ \\ \text{CN} \end{array} \text{---} (\text{CH}_2)_2 \text{---} \text{COO---R---} \begin{array}{c} \text{CH}_3 \\ \\ \text{---Si---O---} \\ \\ \text{CH}_3 \end{array}$ <p>macroinitiateur azoester polysiloxanique</p>

Tableau 2. valeurs adaptées de j et de ψ_s

Intégration		γ	Ψ_0	E
Ψ_1	Ψ_2			
0	1	1,3825	0,5696	$4,5 \cdot 10^{-4}$
0	2	2,3403	0,6670	$2,1 \cdot 10^{-3}$
0	3	2,3602	0,6691	$2,2 \cdot 10^{-3}$
1	2	2,9139	0,7682	$4,3 \cdot 10^{-6}$
1	3	2,8874	0,7647	$1,4 \cdot 10^{-5}$

Le figure 4 pressant K_B et deux ajustements de K_C , en utilisant les deux gammes de l'intégration [0, 2] et [1, 2]. Évidemment l'inclusion des ψ_s dans la gamme de l'intégration affecte la qualité de l'ajustement près du point de commutateur, et pour de plus grandes valeurs. Le Figure 5 pressant la distribution la rate de succès pour $Q = 0,1$, en utilisant K_B et K_c en résolvant l'équation (1).

On peut observe que un des dispositifs les plus attrayants de la conception génétique est qu'il trouve beaucoup de solutions de rechange diverses qui sont très près des valeurs à atteindre désirées de propriété. Le nombre de solutions proche-optimales ou de haut-

forme physique est énuméré dans le figure 4. Le seuil de haut-forme physique est 0,99 pour P1, P3, P4 de conception excepté le polymère 2, qui est 0,985. La conception génétique ne pouvait pas trouver les solutions alternatives avec une valeur de forme physique 0,99 plus grand que pour ce polymère. Tandis que la conception génétique ne trouvait pas les optimums globaux pour le copolymère P 4, elle a localisé plus de 500 à 900 solutions proche-optimales alternatives.

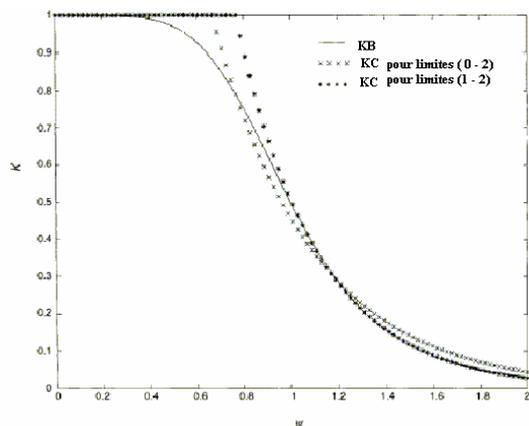


Figure 4.

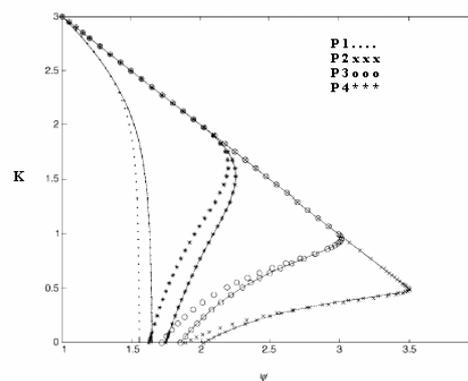


Figure 5.

L'utilisation des méthodes de l'intelligence artificielle dans l'investigation des structures des polymères représente un domaine d'actualité dans le développement de la chimie moderne. En prenant en considération les problèmes majeurs qui apparaissent dans la synthèse des polymères, respectivement la toxicité élevée des monomères, des solvants etc. utilisés par obtenir de nouveaux polymères, mais aussi le grand nombre de produits nécessaires pour trouver un produit final qui réunira certaines propriétés physico-chimiques, on a imposé actuellement l'idée de générer sur ordinateur les conditions de synthèse et des caractéristiques désirées pour les produits finis et de trouver les conditions de synthèse optimale par la comparaison.

CONCLUSIONS

Dans c'est étude nous avons présente l'exécution base au cadre génétique de conception associe a la méthode *Tabou Search* pour réduire l'espaces de recherche. Ces système nous permit également une prolongation du cadre du recherche en incorporant la connaissance chimique de plus haut niveau aux de poignée du contraintes. Ces avances sont démontrées à l'aide d'une étude de cas de conception de polymère. Contrairement à la recherche opérationnelle, l'algorithme génétique n'exige aucune connaissance de la manière dont résoudre le problème; il est seulement nécessaire de pouvoir évaluer la qualité d'une solution. Il est également léger à mettre en œuvre (le «moteur» est commun, il y a moins de programmation spécifique au problème à faire).

On remarque la rate de succès de plus 95% obtenue dans l'analyse réalisée. Evidemment, la complexité des polymères induira la baisse de la rate de succès, mais pour le groupe analysé, les résultats sont encourageants. Les dates obtenues par l'exécution initiale seront introduites comme dates d'apprentissage dans le mécanisme

neuronal, réalisant ainsi l'amélioration de la base d'analyse/comparaison du système. L'adaptabilité spécifique aux réseaux neuronaux permettra ainsi l'exécution des structures plus complexes, partant des prémisses données.

REFERENCES BIBLIOGRAPHIQUES

1. Goldberg, D.E. : *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, **1989**,
2. Glover, F. : Tabu search - part ii, *ORSA Journal on Computing*, **1990**, 2(1), 4-32,
3. Tianzi Jiamg, Cui, O., Shi, G., Ma, S. : *J. Chem. Phys.*, **2003**, 119(8), 4592-4596,
4. Rusu, T., Gogan, O.M. : *Molec. Cryst*, **2004**, 415-418, 155-164,
5. Rusu, T., Bulacovschi, V. : *International Journal of Quantum Chemistry*, 2006, 106(6), 1406-1412,
6. Holland, J.H. : *Adaptation in Natural and Artificial Systems*, The University of Michigan Press, Ann Arbor, **1975**.